長短期記憶神經網路於樂段生成之應用

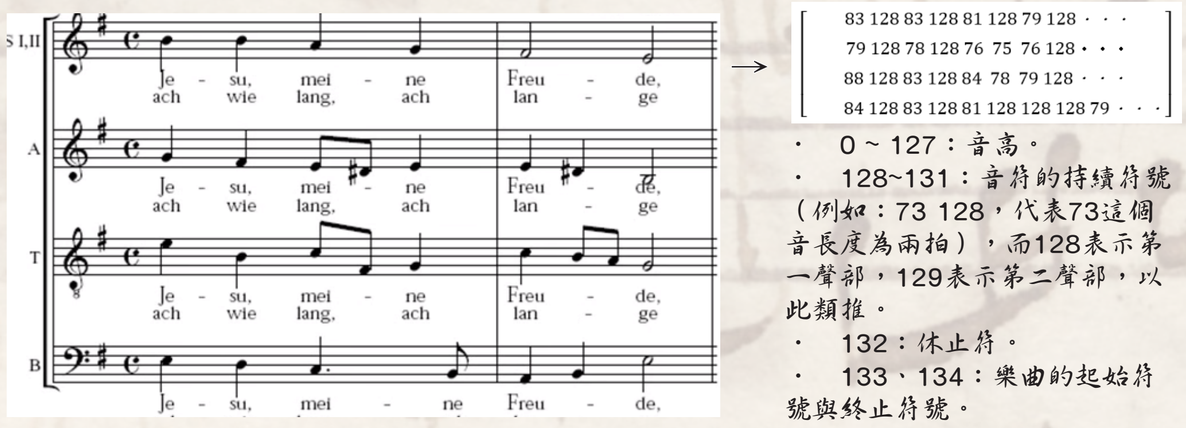
專題編號：107-CSIE-S013
執行期限：106 年第 1 學期至 107 年第 1 學期
指導教授：尤信程 教授
專題參與人員： 104590024 蔡一玄

概述

近年來，人工智慧、機器學習相關領域開始成為科技業的熱門話題。而人工智慧可以應用的領域十分廣泛，只要有一定數量的資料，就能夠讓電腦分析或預測我們想要的結果。在這麼多的領域當中，藝術較難達到比較好的結果，因為當中融入了人類的情感、生命、想法，這些都是電腦難以捉摸的。或許，要創作出具獨特風格以及有靈魂的藝術作品難度甚高，但模仿特定藝術家，創作出類似其風格的作品，這方面是可行的。本專題便是融合音樂以及深度學習相關概念，將電腦可辨識的MIDI檔(一種將音樂以數位方式表達的檔案)進行資料預處理，也就是把音高、拍子．．．等音樂符號轉成矩陣的方式表示，接著利用深度學習中的長短期記憶 (LSTM) 神經網路作為主要模型，產生與輸入矩陣相似的輸出矩陣，再把輸出的矩陣轉回MIDI檔，而這個檔案就是電腦創作的樂曲。

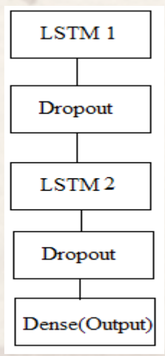
實作方法

首先，將人類看得懂的樂譜轉為電腦可辨識的數字矩陣，每個數字都各自代表不同的訊息。



- 0 ~ 127：音高。
- 128~131：音符的持續符號（例如：73 128，代表73這個音長度為兩拍），而128表示第一聲部，129表示第二聲部，以此類推。
- 132：休止符。
- 133、134：樂曲的起始符號與終止符號。

接著將這些資料分批送入LSTM神經網路模型進行訓練，在訓練之前必須先定義好神經網路的訓練模型，本系統採用最基本的兩個LSTM神經網路作為訓練模型，相對於一個LSTM，兩個LSTM更能夠處理較複雜的輸入資料格式，也能有較好的預測結果。在兩個LSTM神經網路之後我們會接上一層Dropout Layer，透過刪除節點的方式讓神經網路能夠使用不同節點來進行學習，以避免在訓練過程中過度依賴某些節點，而導致產生過擬合(Over-fitting)的問題。輸出層的部分使用的是全連接層(Fully Connected Layer)，會基於前面神經網路所產生的訊號來預測音符。而輸出層的激勵函數採用的是Softmax，可以得知音符的機率分布。



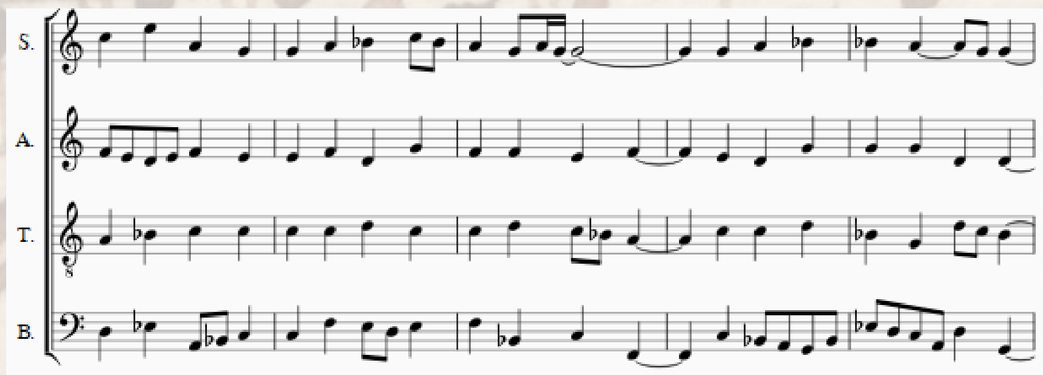
在訓練模型的時候，我們需要提供輸入資料的對應標記，而目標資料(拿來當標記的資料)的格式與輸入資料類似，唯一不同之處是，輸入資料經過轉置後所有聲部的訊息已被混合在一個陣列當中，但在目標資料中，我們為了讓機器分辨某時間點的音符是屬於哪個聲部的，因此將四個聲部的音符拆成四個陣列表示。

訓練完模型之後，我們便可以利用Keras提供的模型預測函式來產生音符的機率矩陣，而透過這個機率矩陣我們可以得知每個時間點的音符機率分布，並取得陣列中機率最高的索引，接著利用這個索引即可解碼出對應的音符或和弦，將它們串接起來後再透過music21所提供的函式轉換為 MIDI檔案，而這個檔案就是電腦生成的音樂。

研究結果

由於標記的方法會直接影響到輸出音樂的結果與品質，因此必須找出一個比較合適的標記方法。

第一種方法是僅將聲部訊息被混合的輸入資料做聲部分類，不做任何時間上的位移，這樣的結果雖然能夠讓神經網路的訓練準確度達到95%以上，但卻導致生成的音樂與輸入音樂相似度非常高。

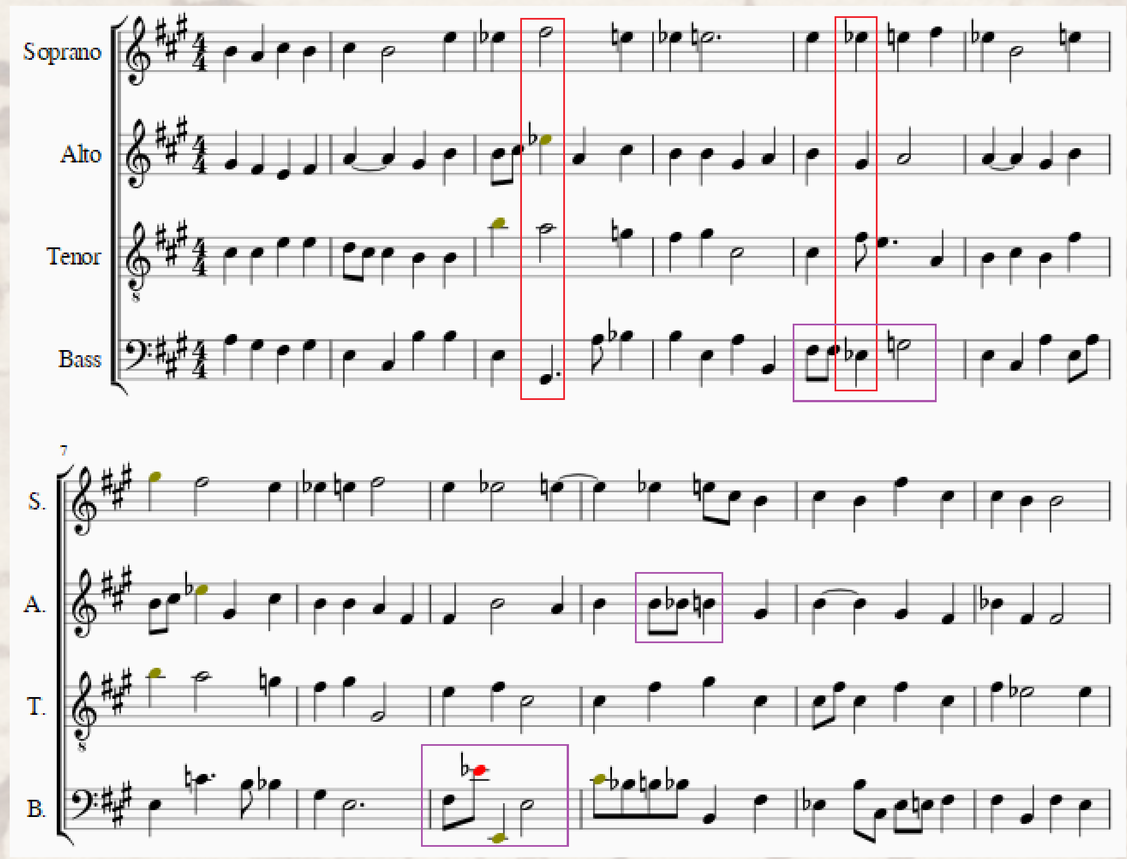


方法一生成的樂段範例

第二種方法是將目標資料做一個小節的位移，也就是當輸入資料為樂曲中的第一小節時，目標資料即為樂曲的第二小節。而以一個小節做位移的原因是，通常一個小樂句或動機的長度差不多是一個小節，這麼做較可以避免產生的音樂中有一些不合常理的節奏。

使用第二種方法產生的音樂樣本中，雖然有一些片段存在不和諧的和聲以及較不規則的旋律配置，但還是能看到一些具有與輸入音樂風格類似又不失獨特性的段落。從下方範例我們可以發現，樂曲當中會不時出現一些較特殊的和聲(縱向方框)，這些和聲可能不適合出現在這個段落，或是違反了一些樂理規則，因此聽起來會有違和感。同樣的，樂曲的片段中也有一些較不和諧的旋律(橫向方框)。

整體而言，方法二生成的樂段雖然有些小缺失，但以創作樂曲的角度而言，這樣的結果會比較接近我們希望達到的樂曲獨特性。



方法二生成的樂段範例